HW3

Arseny Belousov 961152006

# Q1

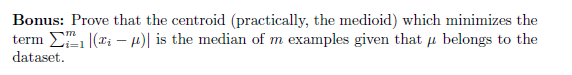
 The K-medoid could be more robust to noise and outliers than the k-means since it minimizes the sum of common paired dissimilarities instead of the sum of the squared Euclidean distances. The possible choice of the dissimilarity function is very rich.

(<http://www.math.le.ac.uk/people/ag153/homepage/KmeansKmedoids/Kmeans_Kmedoids.html#:~:text=In%20contrast%20to%20the%20k,centers%20(%20medoids%20or%20exemplars).&text=It%20could%20be%20more%20robust,sum%20of%20squared%20Euclidean%20distances>.)



We can solve for the mth centroid which minimizes the term by differentiating and setting it equal to 0:

Thus, we can see that the best centroid for minimizing the term is the mean of m examples.



In the same way if we differentiate and set it equal to 0, we can get:

If we solve for xm we find that xm = median, the median of the m examples. The median of a group of points is straightforward to compute and less susceptible to distortion by outliers.

# Q2

(with support of the website <https://towardsdatascience.com/support-vector-machine-simply-explained-fee28eba5496>)

1. Here we have a linear kernel with C=0.01. As we see on figures, we have only two linear options: **A** and **D**. But which one? As we know from lectures, tutorials and see from Google, the bigger C ⬄ the more penalty SVM gets when it makes classification and thus more degree of tolerance. In other words, we see that on the plot A the line slope level is bigger that this one on the plot D => it means that for ***C=0.01*** the relevant plot is the ***plot D***.
2. As the conclusion from the previous exploration, for ***Linear kernel with C=1*** the relevant plot is the ***plot A***.
3. SVM with a polynomial kernel can generate a non-linear decision boundary using polynomial features. We can see that plots with solid nonlinear separations are in Figures C, F and, possibly, E, but the last option is not suitable because the polynomials do not behave as shown in plot E. Thus, there are two options left: C and F. The second order polynomial should be flatter than the 10th order polynomial for obvious reasons: if we consider f (x) = x^2 and f (x) = x^10, for x = (- 2, -1, 0, 1, 2) we will see that an increase in the order leads to a scatter of the values of the function: f(x)=x2: [4,1,0,1,4] and f(x)=x10: [1024,1,0,1,1024]. From that example we can understand that flatter plot is the plot C and thus the ***plot C*** is the relevant for the ***2nd order polynomial kernel***.
4. As we see from the exploration of the previous part, the ***plot F*** must be the one relevant for the ***10th order polynomial kernel***.
5. At this point, we have only two unallocated figures left: figure B and figure E. To understand which figure will correspond to this example, we need to know that the RBF kernel is characterized by a squared Euclidean distance between the two feature vectors. Thus, we can understand that the larger the gamma value, the closer the area border is to the target set of values. Based on the above, we conclude that for the ***RBF kernel with*** the relevant plot is the ***plot E***.
6. As the conclusion from the exploration above and as a logical conclusion we can tell that for the ***RBF kernel with*** the relevant plot is the ***plot B***.

# Q3